

Chapter 3

Is It Worth It? Some Comments on Research and Technology in Assessment and Instruction

J.D. Fletcher
Institute for Defense Analyses

As is true of many things, technology, specifically computer technology, offers both challenges and opportunities. Computer technology is becoming increasingly powerful, ubiquitous, and affordable. Computers are turning up in our automobiles, refrigerators, and hair-dryers, and their effects on our lives and daily routines may have only begun. The challenges this technology presents include rapidly changing work procedures and priorities, which in turn affect what our education and training institutions must do. Computer technology influences not only what we do but also what we choose to do and aspire to accomplish. It affects the structure and organization of our established institutions, as well as the way they go about their business. These issues are as real and challenging for educators concerned with assessment as they are for every other sector of human activity. The effort required to meet these challenges naturally raises questions about whether the promised opportunities outweigh the resources needed to bring them about. In short, is it (the effort) worth it (the new capabilities computer technology offers)? This paper discusses the opportunities and capabilities promised by computer technology for assessing and ensuring human competence, and it suggests some research directions that will help bring these opportunities and capabilities to fruition. It particularly concerns technology used to perform the assessments needed to tailor instruction to the needs of individual students, thereby helping to ensure that the instruction reliably produces its intended outcomes for all. Discussion of these issues, then, may best begin with a perspective on the promise of technology for instruction.

THE THIRD REVOLUTION IN INSTRUCTION

Among other things arising from the ubiquity of computer technology may be a third revolution in instruction—"instruction" being a catch-all term for education, training, and tutoring. From this viewpoint, the first revolution was the development of writing about 7,000 years ago. Writing allowed the content of advanced ideas and instruction to transcend time and place and thereby effect a revolution in instruction. In addition to reviewing trade accounts pressed into mud tablets, people with enough time and resources could study the thoughts of the sages without having to rely on face-to-face interaction or the vagaries of human memory.

The introduction of books produced from moveable type was the second major revolution in instruction. Printed books were first produced in China around 1000 A.D. and in Europe in the mid-1400s (Kilgour, 1998). As with writing, books provided access to learning content that

was available anytime, anywhere, but they also increased accessibility to learning by reducing costs. Books effected major changes in both the techniques and, notably, the objectives of instruction. Curriculum and syllabi were altered to take advantage of the availability of the learning content in books. Moreover, books contributed to the rise of a middle class that, in turn, increased the demand for more access to learning content through more books.

Computer technology may now be effecting a third revolution in instruction. This technology makes both the content and the interactions, the tutorial give-and-take, of learning widely and inexpensively accessible. Computer-based instructional materials are available anytime and anywhere, but they also provide relevant and appropriate instructional interactions. They can be designed to adapt and respond to the needs and intentions of individual learners on a microsecond to microsecond basis. They may foment a third revolution in instruction that is at least as significant as the previous two. We might, therefore, ask if there is any evidence that this revolution is occurring and what role technology-based assessment has played in this activity.

WHAT ARE THE CONTRIBUTIONS OF TECHNOLOGY TO INSTRUCTION?

Computer technology has from the beginning been used interactively to tailor the pace, content, difficulty, and sequencing of instructional material to the needs of individuals. Research, development, use, and assessment of computer applications in instruction began in the mid-1950s. Relevant research and development were well underway by the late 1950s and early 1960s in universities (Holland, 1959; Porter, 1959; Bitzer, Braunfeld, & Lichtenberger, 1962; Suppes, 1964), industry (Uttal, 1962), and the military (Fletcher & Rockway, 1986).

We know that substantial improvements in instructional effectiveness may be obtained by tailoring instruction to the needs and capabilities of individual learners. One widely cited discussion was based on studies performed by Benjamin Bloom and his students (Bloom, 1984), who compared the achievement of individually tutored students (one instructor for each student) with that of classroom students (one instructor for every 28-32 students). It is not surprising to find that individual tutoring in these studies increased the achievement of students. What is surprising is the magnitude of the increase. Bloom reported that the overall difference in achievement across three studies was about two standard deviations, which means, roughly, that tutoring improved the achievement of 50th percentile students to that of 98th percentile students. Two standard deviations is a large difference. Bloom posed it to educators as a 2-sigma challenge.

Why is this 2-sigma difference such a challenge? Why don't we simply provide one-on-one tutoring for all our students? The answer is straightforward and obvious: We can't afford it. The provision of one instructor for each student is, in most cases, prohibitively expensive. Individualized, tutorial instruction seems both an instructional imperative and an economic impossibility.

We may now have the means to break out of this dilemma. Gordon Moore's (famous) law states that the power and memory of computers double about every 18 months (Brenner,

1997). The increasing power and affordability of computer technology, combined with its ability to adapt its interactions in real time and on demand, should help solve the problem for us. Its promise for assessment and instruction has not been lost on researchers and developers.

TECHNOLOGY AND ASSESSMENT IN INSTRUCTION

How might assessment best be used to achieve this promise? One way concerns the speed, or “pace,” at which students learn in classrooms. Classroom teachers regularly report on the differences in the time different students need to achieve instructional objectives. These reports are supported by empirical findings like the following:

- Ratio of time needed by individual kindergarten students to build words from letters: 13 to 1 (Suppes, 1964);
- Ratio of time needed by individual hearing-impaired and Native American students to reach mathematics objectives: 4 to 1 (Suppes, Fletcher, & Zanotti, 1975);
- Overall ratio of time needed by individual students to learn in grades K-8: 5 to 1 (Gettinger, 1984); and
- Ratio of time needed by undergraduates in a major research university to learn features of the LISP programming language: 7 to 1 (private communication, Corbett, 1998).

That these differences exist should come as no surprise. As with Bloom’s findings, what is surprising is their magnitude. Doubtless these differences are due in part to ability, but as Tobias (1982) and others have found, prior knowledge appears to be a major factor, one that quickly overtakes ability in accounting for the speed of learning.

These differences can be accommodated by instruction that takes into account both ability and prior knowledge. Such instruction can take advantage of what students know and concentrate on what they have yet to learn, but tailoring instruction in this way represents a difficult, almost impossible, challenge to classroom teachers working with 20-30 (or more) students. However, technology-based instruction has been tailoring or individualizing instruction practically from its beginning. The benefits of doing so are verified by empirical studies. “Meta-analyses” that compare the time students take to reach a threshold of achievement under technology-based and classroom instruction find an overall time savings of about 30 percent for technology-based instruction (National Research Council [NRC], 1997). These savings matter. For instance, they could reduce by about a fourth the \$4 billion the Department of Defense (DoD) spends annually on specialized skill training.

These savings also matter in our K-12 classrooms. Aside from the obvious motivational issues of keeping students interested and involved in educational material, using their time well will profit both the students and any society that will eventually depend on their competency and achievement. The time-savings offered by technology-based instruction in K-12 education could be more significant and of greater value than those obtained in post-education training.

Often the assessments needed to support this approach are accomplished, even in technology-based instruction, by the use of explicit tests such as we find in Keller’s Personalized

System of Instruction (Keller, 1968). We may now be in a position to progress beyond explicit assessment to something less visible, less obtrusive, and, notably, continuous. Specifically, we may begin to employ the kinds of transparent assessments found in “intelligent” tutoring systems. True systems of this sort are generative—they produce instructional interactions on demand and in real time as needed by individual students. They accomplish this in what has become a commonly accepted practice of maintaining a model of the subject matter, a model of what the student knows or does not know about the subject, and a collection of procedures intended to bring about targeted instructional objectives.

In these applications, the student model is created by analyzing a student’s responses in interactions as they occur and inferring from these what the student knows and does not know by mapping his or her responses onto the “expert” model (represented by the model of the subject matter). Or the student model can consist of a parallel model of the subject matter that accounts for the student’s misconceptions (e.g., Fletcher, 1975; Brown & Burton, 1978; Corbett, Koedinger, & Anderson 1997; VanLehn & Niu, in press). The assessment is accomplished continuously and transparently. This is a promising line of development.

WHAT ARE THE BENEFITS OF ASSESSMENT?

Before investing in such a line of development, we might want to know something about its benefits. Payoffs from assessment transcend instructional applications and extend beyond education to military and industrial applications for screening, classifying, and ranking individuals. These latter applications tend to separate out personnel actions, such as selecting individuals for accession or hiring and classifying them into occupational categories. False positives in these cases can be costly. For example, it costs about \$4 million to fully train an Air Force F-16 pilot and about \$8 million to fully train an F-15 pilot (F15s have two engines and F16s have only one, which accounts for most of this cost difference). It is an expensive matter to select an individual for this type of training if he or she will not be able to complete it successfully.

Aircraft operation is not the only expensive training performed by the military and industry. There are other examples of instruction involving operation, maintenance, and deployment of complex equipment. These costs are increasing because of the continuing infusion of technology into military and industrial operations, and attrition from training is a serious and expensive matter for both sectors. More reliable, valid, and precise assessment to select, classify, and/or certify individuals is at an increasing premium in both sector.

What is the value of our current efforts to select individuals for accession? Within the military, the impact of personnel assessment research has been substantial. Zeidner and Johnson (1989) estimated that savings for the first tour of duty resulting from the Army’s use of personnel selection, classification, and assignment procedures compared to random selection, classification, and assignment are about \$414 million annually and that savings could be increased to \$1 billion annually through simple adjustments in policies and procedures. Improved classification procedures for clerical, surveillance, and communications jobs have been estimated to save the Army \$25 million per year compared to previous methods (Grafton, 1990).

The cost-benefits of some future improvements have also been estimated. An increase of 3 percent in the validity of the current test battery used by the Navy for personnel classification could result in an annual savings of \$83 million in performance improvement (Schmidt, Hunter, & Dunn, 1987). Using the recently developed Enlisted Personnel Allocation System to supplement the current system of classifying soldiers for jobs would save the Army nearly \$480 million per year (Grafton, 1990). The impact of personnel assessment research and development on sectors of the economy outside the military was estimated by Hunter and Schmidt (1982) to be equally substantial. Hunter and Schmidt suggest that the productivity improvement likely to result from replacing univariate selection models with multivariate ones would amount to \$43-54 billion a year. Whatever the actual amounts may be, beneficial results from the continued development and use of personnel assessment procedures on the operational costs of military and civilian organizations are likely.

WHERE DOES TECHNOLOGY COME IN?

How might we improve our personnel assessment procedures? How might we develop precision classification that can identify “aces” for at least some occupation classifications before we begin training or at least very early in the training process? We would like to determine those unique, measurable indicators that characterize a Mozart or a Shakespeare and invest our education and training resources appropriately. Computer technology may make this feasible.

With this technology we may have in hand devices that are capable of opening up and measuring whole new areas of cognition, the significance of which we are now only dimly aware, if at all. More could and should be done to use the unique, multimedia display, timing, and data-recording capabilities of computers to assess knowledge, skills, and abilities of individuals. We may be in a position like that of a person with a telescope not yet turned to the stars or a microscope not yet used to examine a drop of water. We need to look beyond our hard-won, well-wrought psychometric techniques based on paper-and-pencil testing and begin to use our new computer-based tools to full advantage.

Most research and development strategies are built around the concept that scientific principles guide design. This concept is both desirable and feasible, but its opposite is more common. Practice begets principle. We built many bridges before we abstracted bridge-building techniques and principles. In the assessment realm, it may well be time to begin systematic experimentation with many types of new item formats intended to assess the specific, innate capabilities possessed by aces, maestros, and star performers of all sorts. These item formats will produce new conceptions of cognition, which in turn will suggest improved, more targeted item formats. It seems past time to pursue programs intended to promote and encourage such spiral development.

Brown and Burton (1978) embedded such considerations in their “Buggy” computer-assisted instruction program. An entire issue of the *International Journal of Man-Machine Studies* (1982) was devoted to papers on automated psychological testing, many of which involved presentations other than our well-worn multiple-choice items. Hunt and Pellegrino (1984) suggested such an approach as a means to expand our notions of intelligence. A first-rate Air Force laboratory was devoted to exploring these notions until it was disbanded in 1998, when

it was just beginning to document what it was learning about human cognition (e.g., the temporal processing assessment discussed by Chaiken, Kyllonen, & Tirre [2000]). More needs to be done.

ADAPTIVE TESTING

The possibility of adaptive, or “stradadaptive,” testing was studied extensively at the University of Minnesota under a multiyear effort sponsored by the three DoD personnel research and development laboratories and orchestrated by the Office of Naval Research. This work focused on the use of technology to select, in real time, specific multiple-choice test items to be presented to examinees based on their responses to earlier items. Overall, the results of this work showed that tests using adaptive techniques could be shorter, more precise, and reliable (Weiss, 1983).

Adaptive testing might also reduce costs for personnel assessment by using computers to administer and score tests and by requiring fewer test items to accurately assess individuals, but costs were not directly investigated in this effort. Further, only one (Church & Weiss, 1980) of the 16 technical reports produced by this effort concerned the use of non-multiple-choice items and instead investigated items that could only be presented through the unique display capabilities of computers. Nonetheless, adaptive testing using adaptive techniques for presenting and scoring items is a significant advance and has been implemented by the DoD in some high-profile areas. For instance, with more than 270,000 potential recruits taking the Armed Services Vocational Aptitude Battery each year at a cost of about \$20 per administration, the military has a considerable stake in efficient personnel assessment. The Armed Services are now turning to computer technology to provide both the economic benefits of group testing and the precision and flexibility of individual testing. A computerized version of the Armed Services Vocational Ability Test (ASVAB) has been administered to thousands of recruits since 1998. In this case, technology is making an assessment imperative economically feasible.

SIMULATION

Rather than marching individuals through a series of test items, assessments might immerse them in situations like the ones for which they are being selected or prepared. Simulation has been a prominent, long-established technique for both conducting training and assessing the readiness of individuals, crews, teams, groups, and units to perform military operations. Today, it is supported by devices ranging from plastic mock-ups to laptop computers to full-motion aircraft simulators costing more than the aircraft they simulate. Applications range from the operation of oscilloscopes to the repair of computer printers to the deployment of armies. All sectors, educational, industrial, and the military, use techniques ranging from simulated device operation to role-playing in order to prepare and assess personnel. With its current emphasis on “situated learning,” shared mental models, problem solving, and higher-order cognitive processes, instructional use of simulation is becoming as familiar to elementary school children as it is to Air Force pilots and business executives.

But the promise and growth of simulation techniques have masked measurement issues that are now being articulated by psychologists, military commanders, industry leaders, and others who are professionally concerned with assessment. We are just beginning to consider

such psychometric properties of simulation as reliability, validity, and precision, as can be seen in empirical forays into this area by O'Neil and his colleagues (e.g., O'Neil, Allred, & Dennis, 1997a; O'Neil, Chung, & Brown, 1997b). In the free and unscripted flow of simulations, correct decisions can lead to wrong outcomes, and incorrect decisions can lead to success. How do we assess capability under these conditions? Is one pass through a simulation sufficient for assessment or are ten needed? Is one scenario (with its single set of initial conditions) needed or many? Along which dimensions should scenarios be varied? In brief, how should simulated environments be designed to support assessments of individual and group performance?

The realism, or "fidelity," needed by simulations to perform successful assessment is a perennial topic of discussion (e.g., Hays & Singer, 1989; Detterman & Sternberg, 1993). Much of this discussion responds to the intuitive appeal of Thorndike and Woodworth's early argument (1901) for the presence and necessity of identical elements to ensure successful transfer of what is learned in training to what is needed on the job.

Thorndike and Woodworth suggested that such transfer is always specific, never general, and keyed to either substance or procedure. This point of view is echoed in more recent studies of transfer, such as the widely noted paper by Gray and Orasanu (1987) who remark on the "surprising specificity of transfer." As Holding (1991) points out, the identical elements theory is hard to argue with—it seems reasonable to expect task elements mastered in simulation to be performed with some appreciable degree of success on the job.

For dynamic pursuits such as combat where unique situations are frequent and expected, the focus on identical elements often leads to an insistence on maximum fidelity in simulations used for assessment. Because we do not know precisely what will happen, we assume that we must provide as many identical elements as we can. This prescription would suggest a viable approach if fidelity came free, but it does not. As fidelity rises, so do costs. High costs can be borne, but they will also reduce the number, availability, and accessibility of valuable resources that can be routinely provided. We must therefore reduce costs by selecting just the fidelity we need to achieve our objectives. These reductions are as necessary for assessment as they are for training.

There is another issue worth mentioning that involves fidelity, simulation, and assessment. Simulated environments permit an assessment of performance and competence that cannot or should not be attempted without simulation. Aircraft can be crashed, expensive equipment ruined, and lives hazarded in simulated environments in ways that range from impractical to unthinkable without them. Simulated environments provide other benefits for assessment. They can make the invisible visible, compress or expand time, and reproduce events, situations, and decision points over and over. Simulation-based assessment is not a degraded reflection of the real environment we would prefer to use. It allows us to assess aspects of performance that would otherwise be inaccessible.

ASSESSMENT AND NETWORKED SIMULATION

One use of simulation for assessment is receiving increasing and perhaps overdue attention. It concerns the learning and capabilities of collectives (crews, teams, groups, and

organizational units). Concern with collective performance is pervasive and by no means limited to military operations (Cannon-Bowers, Oser, & Flanagan, 1992; Huey & Wickens, 1993). However, in the military, the stakes for collective proficiency are high, and interest in assessing collective behavior is intense. Much current interest in the assessment of collective behavior has centered on the military's development and use of networked simulation.

Networked simulation was originally developed for training applications and was intended to improve the performance of crews, teams, and units (Alluisi, 1991). The individual members of crews, teams, and units who use networked simulation are assumed to be already proficient in their individual skill specialties—they are expected to know how to drive tanks, read maps, fly airplanes, fire weapons, and so on at some acceptable threshold of proficiency before they begin networked simulation exercises. Moreover, the commanders of these crews, teams, and units are expected to possess some basic academic knowledge and practical skills in the command and control of their collectives—they are expected to know at some rudimentary level how to maneuver, use terrain in a tactically appropriate manner, fly helicopters, create and overcome engineered obstacles, etc. The focus in networked simulation is on team rather than individual performance.

Networked simulation consists of modular objects intended to simulate combat entities. Typical entities are vehicles such as tanks, helicopters, and aircraft. During simulation exercises, these vehicles are mostly operated by human crews located in the devices that simulate them. These entities, these simulators, may be located anywhere because they are modular and autonomous and because they all share a common model of the battlefield and its terrain. In a networked simulation exercise conducted on simulated California terrain, a tank crew sitting in a simulated tank in Germany can call for air support from simulated aircraft in Nevada because they are being attacked by a simulated helicopter located in Alabama.

Each entity, along with many others, is connected to the network. If the simulated vehicles encounter allied vehicles on the digital terrain, they can join together to form a larger team and undertake a mission with all the problems of command, control, communications, coordination, timing, and so on that such activity presents. If they encounter enemy vehicles, they can engage in force-on-force engagements in which the outcome is determined solely by the performance of the individuals, crews, teams, and units involved. No umpires, battlemasters, or other outside influences are expected or permitted to affect the outcome of a networked simulation engagement once it begins.

All the digital communication packets used to control networked simulation may be recorded. Generally, each entity issues 3-5 packets per second. Actions undertaken in networked simulation may be recorded in extensive detail for later analyses and replay during After Action Reviews (Meliza, Bessemer, & Hiller, 1994; Morrison & Meliza, 1999). The scene from any vantage point (friendly or enemy, inside or outside vehicles, ground level or "God's eye") can be recorded at almost any level of detail and then replayed for the purposes of assessment. Packets have even been created and used to replay entire battles, such as the 73 Easting combat engagement during the Gulf War (Orlansky & Thorpe, 1992).

Use of networked simulation in assessment has been discussed by Fletcher (1994, 1999) and O'Neil et al. (1997b). The paper by O'Neil and his colleagues is particularly interesting because of its presentation of empirical data on the validity of networked simulation used to assess performance on negotiation tasks. Empirical evaluations concerning the training value of networked simulation used by the military have been summarized by Fletcher (1999) and Orlansky, Taylor, Levine, & Honig (1997).

The report by Orlansky et al. is notable for its careful examination of the cost benefits of networked simulation. These researchers compared the costs of a 5-day close air support (aircraft and ground forces operating together) exercise using linked simulators located in Arizona, Kentucky, and Maryland with a "live" simulation performed in the field using actual equipment. The simulation exercise involved 75 people; a similar exercise in the field with actual equipment would have required 245 people. It cost \$267,000 to support the simulation exercise; the field exercise would have cost \$2,897,000. Cost per person trained and assessed in the simulation exercise was \$3,600; cost per person trained in the field would have been \$11,800. As is typical for combat exercises, it was not possible to validate the results of the exercise with real experience (a situation for which we may all be grateful), but steady improvements in combat-relevant tasks were found in the simulation exercise, and its cost benefits for both training and assessment were clearly evident.

Civilian applications of networked simulation for training and education were identified and discussed by Fitzsimmons and Fletcher (1995). These applications were both potential and real. They included two demonstrations involving high school students in DoD schools in Germany, Kentucky, and Korea who collaborated in playing music together ("The World Band") and in designing and flying aircraft using materials available in the early 1900s ("The Wright Flyer"). Although the emphasis in these demonstrations was on education, assessment of such collective issues as teamwork, communication, leadership, interpersonal skills, etc., could easily have been carried out in these demonstrations.

WHERE ARE WE HEADED?

When we consider the possibilities for the use of technology in assessment, it seems reasonable to ask, what will be next? Technology-based instruction appears to be headed for distributed (anytime, anywhere) lifelong learning. It may even be object-oriented, using instructional objects available on the World Wide Web or whatever the global ether will be in the future. These objects will be assembled, on-demand, in real time, in some granular, perhaps item-by-item basis, and tailored to the needs, capabilities, and intentions of individual users, who may be learners, users seeking decision aids, or individuals needing certification for some set of knowledge and skills. The challenges presented by this future are being addressed by the Advanced Distributed Learning (ADL) initiative, which is led by the Department of Defense in coordination with other federal agencies such as the Departments of Agriculture, Education, Labor, Interior, and Health and Human Services; National Aeronautics and Space Administration; National Institute for Standards and Technology; and the White House Office of Science and Technology Policy (<http://www.adlnet.org>).

The Department of Defense is coordinating development with industry of a Sharable Content Objects Reference Model (SCORM) to ensure accessibility, durability, portability, and reusability of instructional objects and to provide guidelines concerning the creation, archiving, and assembly of instructional objects into relevant instructional presentations. Benefits in terms of saved or avoided personnel and training costs are very close to those identified for technology-based instruction (discussed earlier in this paper). Benefits in terms of improved productivity and effectiveness are more difficult to assess, but they are expected to exceed the monetary value of the ADL initiative. The benefits of allowing assessment to take place at any time, any place, and as needed seem likely but have yet to be systematically determined. That such assessment capabilities will be developed seems equally likely.

In any case, assessment can take advantage of sharable objects. Much, however, remains to be done. How, for instance, can we assemble, aggregate, and sequence different objects at different times to produce assessments that are both fair and comprehensive? Should psychometric data be included in the “meta-data” in which objects are packaged? What do we need to do to certify the quality of these objects? These questions, among others, remain as challenges to those who are concerned with what might be described as object-oriented, technology-based assessment.

FINAL WORD

The above comments suggest a number of areas for research. Four that might be emphasized here are:

- *Transparent, continuous assessment.* How do we, or should we, extract assessment information from the interactions between a student and a teacher, whether human or computer? Master teachers know some of the techniques for doing this, and others have been developed for intelligent tutoring systems. More could and should be done. Our current processes of extracting assessment information once every few years, once a year, or even once a month are insufficient if we hope to use instructional and student time well. The hallmark of good management is continuous assessment. We should develop it.
- *Precision classification.* Every human being should have the assessment tools to develop to its fullest extent whatever package of abilities he or she has been handed at birth. We need more comprehensive models of cognition to do this. These models will have to be keyed to our ability to measure them. Through computer technology, we may have in hand the capabilities to devise new item formats and to pursue, in a spiral of development, both the measures and the models of cognition we need. It seems past time to begin this work in earnest.
- *Assessment based on simulation.* Simulation is widely used by industry and the military to assess the capabilities and preparation of individuals, crews, teams, and units. Given the current emphasis (which despite its rhetorical fluff seems sensible) on approaches involving situated, problem- or project-based learning in (more or less) authentic environments—which are very close to, if not the same thing as, what the military calls simulations—the need to determine what students are learning from these simulated environments seems likely to grow. But how many simulations using what scenarios are

needed to ensure reliable, valid, fair assessment? What are the measurement properties of simulations, and how should we develop them further? There is a great need in both education and military and industrial training for answers to these questions—answers that again must come from vigorous, targeted programs of instruction

- *Object-oriented assessment.* The vision of a World Wide Web heavily populated with objects that are accessible, portable, durable, and reusable seems very likely to occur. These objects are likely to include assessment as well as instructional objects. How should we use these objects to assemble assessments in real time and on-demand as needed by individuals? How would we develop the measurement properties of such presentations to ensure reliability, validity, and fairness? Given the advances made by such efforts as the ADL initiative, we are in a good position to begin the necessary research and development. Again, it seems the time is ripe to begin doing so.

All of these areas present challenges to assessment. As suggested, technology will change not only the way we do assessment but our objectives and expectations for assessment as well. The object of assessment is, of course, not better measurement, although that is clearly an enabling capability. What we seek are better (more reliable, valid, and precise) inferences and decisions based on our assessment. Technology will allow access to areas of human cognition and performance we have been unable to consider with our paper-based techniques, and this, in turn, will necessitate new notions of human cognition and potential. It may enable us to identify human capabilities that might otherwise remain latent and undeveloped. The challenges presented include great opportunities.

In the area of human cognition, we may well seek to identify something that might be called (and has been so called by CRESST) a “learnome.” The human genome lists all the micro-components needed for reproduction or replication; the learnome might list all the micro-components needed to reproduce or replicate areas of knowledge or skills. First we need to identify—and measure—these components. If we are successful, we will have made significant progress toward new concepts of cognition and our ability to assess performance of very complex tasks, which seem to be growing increasingly common in both industry and the military (NRC, 1997).

Finally, e-learning is increasing emphasis on learner, as opposed to teacher, classroom, or school, productivity. Learners are expected to be self-motivated, self-guided, and self-regulating in the Webbed world of lifelong learning. Such activity benefits the individual seeking to achieve his or her potential, the organizations depending for their success on human competence, and the nations competing in the global marketplace. All these ends are likely to be well served by tools placed in learners’ hands to help them assess progress toward their goals. Technology seems key in developing these assessment tools and making them available anytime and anywhere to those who need them.

REFERENCES

- Alluisi, E.A. (1991). The development of technology for collective training: SIMNET, a case history. *Human Factors*, 33, 343-362.
- Bitzer, D.L., Braunfeld, P.G., & Lichtenberger, W.W. (1962). Plato II: A multiple-student, computer-controlled, automatic teaching device. In J.E. Coulson (Ed.), *Programmed learning and computer-based instruction* (pp. 205-216). New York: John Wiley.
- Bloom, B.S. (1984). The 2-sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13(6), 4-16.
- Brenner, A.E. (1997). Moore's law. *Science*, 275, 1551.
- Brown, J.S., & Burton, R.B. (1978). Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive Science*, 2, 155-192.
- Cannon-Bowers, J.A., Oser, R., & Flanagan, D.L. (1992). Work teams in industry: A selected review and proposed framework. In R.W. Swezey & E. Salas (Eds.), *Teams: Their training and performance* (pp. 355-377). Norwood, NJ: Ablex.
- Chaiken, S.R., Kyllonen, P.C., & Tirre, W.C. (2000). Organization and components of psychomotorability. *Cognitive Psychology*, 40, 198-226.
- Church, A.T., & Weiss, D.J. (1980). *Interactive computer administration of a spatial reasoning test* (Research Report 80-2). Minneapolis, MN: Computerized Adaptive Testing Laboratory, University of Minnesota.
- Corbett, A.T., Koedinger, K.R., & Anderson, J.R. (1997). Intelligent tutoring systems. In M.G. Helander, T.K. Landauer, & P.V. Prabhu (Eds.), *Handbook of human-computer interaction* (pp. 849-874). Amsterdam: Elsevier Science.
- Detterman, D.K., & Sternberg, R.J. (Eds.). (1993). *Transfer on trial: Intelligence, cognition, and instruction*. Norwood, NJ: Ablex.
- Fitzsimmons, E.A., & Fletcher, J.D. (1995). Beyond DoD: Non-Defense training and education applications of DIS. *Proceedings of the IEEE*, 83, 1179-1187.
- Fletcher, J.D., & Rockway, M.R. (1986). Computer-based training in the military. In J.A. Ellis (Ed.), *Military contributions to instructional technology* (pp. 171-222). New York: Praeger.
- Fletcher, J.D. (1975). Modeling the learner in computer-assisted instruction. *Journal of Computer-Based Instruction*, 1, 118-126.
- Fletcher, J.D. (1994). What networked simulation offers to the assessment of collectives. In H.F. O'Neil, Jr. & E.L. Baker (Eds.), *Technology assessment in software applications* (pp. 255-272). Hillsdale, NJ: Lawrence Erlbaum.
- Fletcher, J.D. (1999). Using networked simulation to assess problem solving by tactical teams. *Computers in Human Behavior*, 15, 375-402.
- Gettinger, M. (1984). Individual differences in time needed for learning: A review of the literature. *Educational Psychologist*, 19, 15-29.
- Grafton, F. (1990). Improving the selection, classification, and utilization of Army enlisted personnel. In J. Orlansky, F. Grafton, C.J. Martin, W. Alley, & B. Bloxom (Eds.), *The*

- current status of research and development on selection and classification of enlisted personnel* (IDA Document D-715). Alexandria, VA: Institute for Defense Analyses.
- Gray, W.D., & Orasanu, J.M. (1987). Transfer of cognitive skills. In S.M. Cormier & J.D. Hagman (Eds.), *Transfer of learning* (pp. 183-215). New York: Academic Press.
- Hays, R.T., & Singer, M.J. (1989). *Simulation fidelity in training system design: Bridging the gap between reality and training*. New York: Springer-Verlag.
- Holding, D.H. (1991). Transfer of training. In J.E. Morrison (Ed.), *Training for performance: Principles of applied human learning* (pp. 93-125). New York: John Wiley.
- Holland, J. (1959). A teaching machine program in psychology. In E. Galanter (Ed.), *Automatic teaching: The state of the art* (pp. 69-82). New York: John Wiley.
- Huey, B.M., & Wickens, C.D. (1993). *Workload transition: Implications for individual and team performance*. Washington, DC: National Academy Press.
- Hunt, E., & Pellegrino, J. (1984). *Using interactive computing to expand intelligence testing. A critique and prospectus* (Report 84-2).
- Hunter, J., & Schmidt, F. (1982). Fitting people to jobs: The impact of personnel selection on national productivity. In M.D. Dunnette & E.A. Fleishman (Eds.), *Human performance and productivity: Human capability assessment*. Hillsdale, NJ: Lawrence Erlbaum.
- Keller, F.S. (1968). Goodbye, teacher *Journal of Applied Behavior Analysis*, 1, 79-89.
- Kilgour, F.G. (1998). *The Evolution of the Book*. New York, NY: Oxford University Press.
- Meliza, L.L., Bessemer, D.W., & Hiller, J.A. (1994). Providing unit training feedback in the distributed interactive simulation environment. In R.F. Holtz, J.A. Hiller, & H.H. McFann (Eds.), *Determinants of effective unit performance* (pp. 257-280). Alexandria, VA: U.S. Army Research Institute.
- Morrison, J.E., & Meliza, L.L. (1999). *Foundations of the After Action Review Process* (IDA Document 2332). Alexandria, VA: Institute for Defense Analyses. (DTIC/NTIS AD-A368 651)
- National Research Council (1997). *Technology for the United States Navy and Marine Corps, 2000-2035: Becoming a 21st century force vol. 4 Human Resources*. Committee on Technology for Future National Forces. Washington, DC: National Academy Press.
- O'Neil, H.F., Allred, K., & Dennis, R.A. (1997). Validation of a computer simulation for assessment of interpersonal skill. In H.F. O'Neil (Ed.), *Workplace readiness: Competencies and assessment* (pp. 229-254). Mahwah, NJ: Lawrence Erlbaum.
- O'Neil, H.F., Chung, G.K.W.K., & Brown, R.S. (1997). Use of networked simulations as a context to measure team competencies. In H.F. O'Neil (Ed.), *Workplace readiness: Competencies and assessment* (pp. 411-452). Mahwah, NJ: Lawrence Erlbaum.
- Orlansky, J., Taylor, H.L., Levine, D.B., & Honig, J.G. (1997). *The cost and effectiveness of the Multi-Service Distributed Training Testbed (MDT2) for training close air support* (IDA Paper P-3284). Alexandria, VA: Institute for Defense Analyses.

- Orlanksy, J., & Thorpe, J. (1992). *Proceedings of conference on 73 Easting: Lessons from Desert Storm via advanced simulation technology held in Alexandria, Virginia on 27-29 August 1991* (IDA Document 1110). Alexandria, VA: Institute for Defense Analyses. (DTIC/NTIS AD-A253 991)
- Porter, D. (1959). Some effects of year long teaching machine instruction. In E. Galanter (Ed.), *Automatic teaching: The state of the art* (pp. 85-90). New York: John Wiley.
- Schmidt, F.L., Hunter, J.E., & Dunn, W.L. (1987). *Potential utility increases from adding new tests to the Armed Services Vocational Aptitude Battery (ASVAB)*. Unpublished manuscript (Battelle Contract Delivery Order 53). San Diego, CA: Navy Personnel Research and Development Center.
- Suppes, P. (1964). Modern learning theory and the elementary-school curriculum. *American Educational Research Journal*, 1, 79-93.
- Suppes, P., Fletcher, J.D., & Zanotti, M. (1975). Performance models of American Indian students on computer-assisted instruction in elementary mathematics. *Instructional Science*, 4, 303-313.
- Thorndike, E.L., & Woodworth, R.S. (1901). The influence of improvement in one mental function upon the efficiency of other functions. *Psychological Review*, 8, 247-262.
- Tobias, S. (1982). When do instructional methods make a difference? *Educational Researcher*, 11(4), 4-9.
- Uttal, W.R. (1962). On conversational interaction. In J.E. Coulson (Ed.), *Programmed learning and computer-based instruction* (pp. 171-190). New York: John Wiley.
- VanLehn, K., & Niu, Z. (in press). Bayesian student modeling, user interfaces and feedback: A sensitivity analysis. *International Journal of Artificial Intelligence in Education*, 12.
- Weiss, D.J. (1983). *Final report: Computer-based measurement of intellectual capabilities*. Minneapolis, MN: Computerized Adaptive Testing Laboratory, University of Minnesota.
- Zeidner, J., & Johnson, C.D. (1989). *The economic benefits of predicting job performance* (IDA Paper P-2241). Alexandria, VA: Institute for Defense Analyses.